Distribusi Peluang

Pendahuluan

Pokok bahasan yang akan Anda pelajari dalam modul ini adalah distribusi peluang dan sifat-sifatnya. Pokok bahasan ini terdiri dari tiga subpokok bahasan, yaitu distribusi peluang, beberapa sifat distribusi peluang, dan distribusi peluang bersama.

Setelah mempelajari modul ini, secara umum Anda diharapkan dapat memahami distribusi peluang dan sifat-sifatnya.

Secara khusus, Anda diharapkan mampu:

1. menentukan distribusi peluang dari suatu variabel random,
2. menghitung nilai harapan dan variansi dari variabel random,
3. menggunakan berbagai sifat mean dan variansi dari variabel random,
4. menghitung distribusi peluang bersama dari dua variabel random,

Kegiatan Belajar

1
Distribusi Peluang

Dalam suatu eksperimen, kita sering menjumpai hasil elementer yang merupakan deskripsi kualitatif, bukan nilai numerik. Misalnya, dalam pelemparan satu mata uang logam dua kali, hasil-hasil elementernya adalah *MM*, *MB*, *BM*, *BB* yang menunjukkan hasil suatu lemparan muka *(M)* atau belakang *(B)*. Jika akibat sampingan suatu obat baru dipelajari, apakah mengakibatkan mual atau tidak, mungkin jawabannya sangat mual, agak mual, tidak mual sama sekali. Ini adalah hasil kualitatif, bukan hasil pengukuran dengan skala numerik.

Sering kali pula kita jumpai hasil suatu eksperimen adalah nilai numerik. Misalnya, banyak kecelakaan lalu-lintas setiap hari di suatu kota, upah harian pekerja pabrik, nilai setiap calon yang menempuh UMPTN. Dalam hal hasil elementer berbentuk deskripsi kualitatif, sering kali kita juga ingin memperoleh gambaran numerik dari hasil kualitatif itu. Misalnya, dari percobaan obat baru untuk 100 orang, informasi untuk menilai obat baru itu dapat berbentuk berapa orang merasa sangat mual, agak mual, atau tidak mual sama sekali. Demikian juga, hasil survei pendapat bagi 500 kepala keluarga di suatu kota tentang masalah yang timbul di kota itu, informasi yang diperoleh dapat berbentuk berapa banyak yang mendukung dan berapa pula yang menentang. Dalam contoh-contoh ini, observasi individual tidak berbentuk numerik, tetapi ringkasan numerik himpunan observasi itu dapat digunakan sebagai dasar inferensi. Dalam modul ini, kita akan memusatkan perhatian kita pada aspek numerik hasil-hasil eksperimen. Untuk itu kita kenalkan konsep *variabel random*.

Variabel random *X* adalah cara memberi nilai angka pada setiap hasil suatu eksperimen yang ditentukan dari sifat yang ada pada hasil itu. Dalam bahasa matematika kita katakan bahwa variabel random adalah fungsi bernilai real yang didefinisikan pada suatu ruang sampel. Perkataan *random* merupakan peringatan bahwa sebelum eksperimen kita tidak tahu hasil eksperimen itu atau pun nilai variabel *X* yang berkaitan dengannya.

Contoh 1.1

Jika X menyatakan banyak muka (M) yang diperoleh dalam tiga pelemparan satu mata uang logam, maka X adalah variabel karena banyak muka dalam tiga pelemparan satu mata uang logam dapat mempunyai salah satu nilai 0, 1, 2, atau 3. Variabel ini random dalam arti bahwa nilai yang akan terjadi dalam tiga pelemparan itu tidak dapat diperkirakan dengan pasti. Namun demikian, kita dapat membuat tabel hasil elementer dan nilai X yang berkaitan dengannya sebagai berikut.

Hasil	**Nilai *X***
BBB	0
BBM	1
BMB	1
MBB	1
BMM	2
MBM	2
MMB	2
MMM	3

Perhatikan bahwa untuk setiap hasil elementer terdapat satu nilai, tetapi beberapa hasil elementer dapat mempunyai nilai yang sama. Selanjutnya, dengan mencermati tabel di atas, kita dapat mengelompokkan peristiwa-peristiwa (yakni himpunan hasil elementer) yang berkaitan dengan nilai-nilai *X* yang berbeda-beda dalam tabel berikut.

Nilai Numerik *X*	**Komposisi Peristiwa**
(X = 0)	{*BBB*}
(X = 1)	{*MBB, BMB, BBM*}
(X = 2)	{*MMB, MBM, BMM*}
(X = 3)	{*MMM*}

Dari Contoh 1.1, kita amati fakta umum bahwa peristiwa-peristiwa yang berkaitan dengan nilai-nilai X yang berbeda tidak mempunyai elemen-elemen yang berserikat. Karena itu gabungan peristiwa-peristiwa ini merupakan ruang sampelnya. Biasanya, nilai-nilai variabel random X yang mungkin dapat ditentukan secara langsung dari deskripsi variabel random itu tanpa membuat daftar elemen-elemen ruang sampelnya. Tetapi, untuk menetapkan nilai peluang bagi nilai-nilai variabel ini, yang kita pandang sebagai peristiwa, sering kali akan lebih enak jika kita kembali ke ruang sampelnya.

Contoh 1.2

Sekumpulan 12 juri diminta untuk membandingkan rasa makanan A yang dibuat oleh juru masak wanita dan juru masak pria. Misalkan X menunjukkan banyak juri yang menilai masakan wanita setidaknya sama enak dengan masakan pria. Di sini X dapat menjalani nilai-nilai 0, 1, 2, …, 12.

Contoh 1.3

Pada perempatan jalan, seorang pengamat akan menghitung banyak mobil (X) yang lewat sampai satu mobil dengan nomor polisi baru lewat, maka nilai-nilai yang mungkin dijalani X adalah 1, 2, 3, … (teoretis barisan bilangan itu dapat tidak berhenti).

Suatu variabel random dikatakan *diskret* jika variabel itu hanya menjalani nilai-nilai yang banyaknya berhingga atau nilai-nilai yang tak berhingga banyak yang dapat disusun dalam barisan. Semua variabel random dalam contoh-contoh tersebut adalah *variabel random diskret*. Sebaliknya, jika suatu variabel random itu merupakan hasil pengukuran dengan skala kontinu sehingga dapat menjalani semua nilai-nilai dalam suatu interval, maka dinamakan *variabel random kontinu.* Tentu saja setiap alat pengukur mempunyai ketepatan yang terbatas, dengan demikian skala kontinu harus diinterpretasikan sebagai suatu abstraksi. Beberapa contoh variabel random kontinu adalah tinggi badan orang dewasa, hasil susu per hari suatu perusahaan pemerahan susu, lama hidup seorang pasien setelah menjalani serangan jantung.

Dalam modul ini akan kita bicarakan distribusi peluang variabel random diskret. Sedangkan distribusi peluang variabel random kontinu akan kita bicarakan dalam modul-modul berikutnya. Dalam proses konseptualisasi, distribusi peluang variabel random kontinu ini terlibat pandangan/pemikiran yang cukup berbeda dengan distribusi peluang variabel random diskret.

Distribusi Peluang Variabel Random Diskret

Daftar nilai-nilai yang mungkin dari suatu variabel random X membuat kita sadar akan semua hasil yang mungkin dari suatu eksperimen. Dengan menggunakan konsep peluang, kita dapat menentukan peluang akan mengamati berbagai nilai. Untuk ini kita kenalkan pengertian tentang distribusi peluang (fungsi peluang).

Distribusi peluang atau dengan singkat disebut *distribusi* suatu variabel random diskret X adalah daftar nilai-nilai numerik X yang berbeda bersama dengan nilai peluangnya. Kadang-kadang rumus matematik dapat digunakan untuk merepresentasikan distribusi peluang ini.

Contoh 1.4

Jika X menunjukkan banyak muka yang diperoleh dalam tiga lemparan sebuah mata uang logam yang seimbang, maka distribusi peluang dari X dapat dihitung sebagai berikut.

Dalam Contoh 1.1 telah dibuat tabel delapan hasil elementer dan nilai X yang berkaitan dengannya. Nilai-nilai X yang berbeda adalah 0, 1, 2, dan 3. Sekarang kita hitung peluang masing-masing. Model untuk mata uang yang seimbang memberikan delapan hasil elementer berkemungkinan sama, sehingga untuk masing-masing ditetapkan peluang $\frac{1}{8}$. Peristiwa $(X=0)$ mempunyai satu hasil yaitu *BBB*, sehingga peluangnya $\frac{1}{8}$. Demikian juga, peluang $(X=1)$, $(X=2)$, dan $(X=3)$ masing-masing adalah $\frac{3}{8}$, $\frac{3}{8}$, dan $\frac{3}{8}$. Hasil-hasil ini dikumpulkan dan diperoleh distribusi peluang seperti dalam Tabel 1.1.

Tabel 1.1. Distribusi Peluang

Nilai *X*	**Peluang**
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{3}{8}$
Jumlah	1

Dalam pembicaraan umum, kita akan menggunakan notasi x_1, x_2 dan seterusnya untuk menunjukkan nilai-nilai yang berbeda variabel random X. Peluang bahwa satu nilai tertentu x_i terjadi akan ditulis dengan $f(x_i)$. Seperti dalam Contoh 1.1, jika X

menjalani k nilai-nilai yang mungkin yaitu $x_1, x_2, \ldots, x_k$ dengan nilai peluang masing-masing $f(x_i)$, $f(x_2)$, ..., $f(x_k)$ maka distribusi peluang dari X dapat disajikan dalam Tabel 1.2. Karena kuantitas $f(x_i)$ menunjukkan nilai peluang, maka $f(x_i)$ harus merupakan bilangan antara 0 dan 1. Dan lebih dari itu, jika kita jumlahkan untuk semua nilai X yang mungkin harus sama dengan 1. Distribusi peluang atau fungsi peluang menggambarkan bagaimana total peluang 1 terbagi-bagi untuk nilai-nilai random tersebut.

Tabel 1.2. Bentuk Distribusi Peluang Diskret

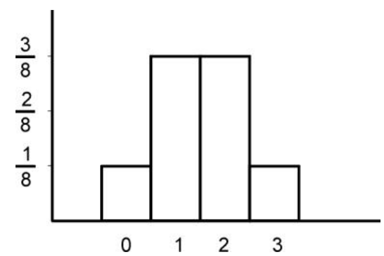
Nilai *x*	**Peluang** $f(x_i)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_k	$f(x_k)$
Jumlah	1

Distribusi peluang suatu variabel random diskret X digambarkan sebagai fungsi:

$$f(x_i) = P(X = x_i)$$

dan mempunyai sifat $f(x_i) \geq 0$ untuk setiap x_i dan $\sum_{i=1}^{k} f(x_i) = 1$.

Dengan penyajian grafik, suatu distribusi peluang dapat membantu mengungkapkan adanya pola dalam distribusi peluang itu. Penyajian dalam bentuk histogram frekuensi juga akan memberikan jalan ke arah pembentukan konsep distribusi kontinu. Untuk menggambarkan distribusi peluang, pertama-tama kita tuangkan nilai-nilai X pada sumbu mendatar. Dengan tiap-tiap nilai x_i sebagai pusat, kita buat empat persegi panjang tegak yang luasnya sama dengan peluang $f(x_i)$. Histogram peluang untuk distribusi dalam Contoh 1.4 ditunjukkan dalam Gambar 1.1 berikut.



Gambar 1.1. Histogram Peluang *X*

Contoh 1.5

Misalkan 30% dari pohon-pohon di suatu hutan terserang parasit. Empat pohon dipilih secara random. Jika X menunjukkan banyak pohon dalam sampel yang terserang parasit, maka distribusi peluang X dapat dihitung dan histogran peluangnya dapat digambarkan sebagai berikut.

Karena tiap pohon dapat terserang *(S)* atau tidak terserang *(T)* parasit maka banyak hasil elementer dalam sampel dengan empat pohon adalah $2 \times 2 \times 2 \times 2 = 16$, sehingga diperoleh daftar menurut nilai X sebagai berikut.

$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$
TTTT	*TTTS*	*TTSS*	*TSSS*	*SSSS*
	TTST	*TSTS*	*STSS*	
	TSTT	*TSST*	*SSTS*	
	STTT	*STTS*	*SSST*	
		STST		
		SSTT		

Selanjutnya, peluang untuk tiap-tiap nilai X dapat dihitung sebagai berikut. Pertama-tama, kita perhatikan penetapan nilai peluang bagi hasil elementernya. Untuk satu pohon yang dipilih secara random kita punyai $P(S)=0,3$ dan $P(T)=0,7$ karena 30% dari populasi pohon-pohon itu terserang parasit. Lagi pula, populasinya sangat besar (hutan memuat sangat banyak pohon) dan ukuran sampelnya sangat kecil, maka praktis observasi empat pohon itu dapat dipandang sebagai independen. Dengan menggunakan sifat independensi dan memberlakukan hukum perkalian peluang, kita hitung:

$$P(X=0)=P(TTTT)=(0,7)(0,7)(0,7)(0,7)=0,2401$$

Peristiwa $(X=1)$ mempunyai 4 hasil elementer, di mana masing-masing memuat tiga T dan satu S dengan

$$P(TTTS)=(0,7)(0,7)(0,7)(0,3)=0,1029$$

maka kita peroleh $P(X=1)=(4)(0,1029)=0,4116$

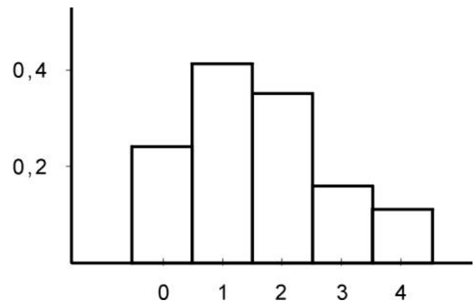
Selanjutnya, dengan cara yang sama diperoleh:

$$P(X=2)=(6)(0,7)^2(0,3)^2=0,2646$$
$$P(X=3)=(4)(0,7)(0,3)^3=0,0756$$
$$P(X=4)=(0,3)^4=0,0081$$

Dengan mengumpulkan hasil-hasil ini, distribusi peluang dari X kita sajikan dalam Tabel 1.3, dan histogram peluangnya kita tuangkan dalam Gambar 1 2.

Tabel 1.3. Distribusi Peluang *X*

x	$f(x)$
0	0,2401
1	0,4116
2	0,2646
3	0,0756
4	0,0081
Jumlah	1,0000



Gambar 1.2. Histogram Peluang

Dalam kesempatan ini, kita akan menjelaskan dengan singkat peranan distribusi peluang dalam inferensi statistik. Guna menghitung peluang yang berkaitan dengan nilai-nilai suatu variabel random, kita perlu pengetahuan penuh tentang ketidakpastian hasil-hasil eksperimen. Sebagai contoh, jika *X* menunjukkan sesuatu sifat numerik sampel random dari suatu populasi, kita menganggap komposisi dari populasi itu diketahui supaya distribusi dari *X* dapat dihitung secara numerik. Dalam Contoh 1.5, peluang akan mengamati berbagai nilai *X* dihitung dengan anggapan bahwa tingkat terjadinya serangan parasit dalam populasi pohon-pohon itu adalah 0,3. Dalam aplikasi praktis, biasanya kuantitas proporsi (parameter) itu tidak diketahui. Misalnya huruf *p* kita gunakan untuk menunjukkan proporsi pohon-pohon yang terserang parasit (yang tidak diketahui). Inferensi statistik mencoba untuk menentukan nilai *p* yang dipandang dapat diterima berdasarkan nilai-nilai observasi *X* dalam sampel. Untuk memantapkan pengertian inferensi, misalkan keempat pohon dalam sampel terserang parasit, apakah 0,3 merupakan nilai *p* yang dapat diterima? Tabel 1.3 menunjukkan bahwa jika *p* benar-benar sama dengan 0,3 maka peluang akan mengamati nilai $X = 4$ hanyalah 0,0081. Nilai peluang yang sangat rendah ini membuat keraguan pada hipotesis bahwa $p = 0,3$. Penalaran statistik seperti ini akan kita pelajari lebih jauh dalam pembahasan berikutnya.

Distribusi peluang dalam Contoh 1.4 dan Contoh 1.5 diperoleh dengan pertama-tama menetapkan peluang hasil elementer dengan menggunakan proses deduksi logis. Jika ini tidak dapat dilakukan, kita harus menggunakan cara penentuan empiris. Ini melibatkan pengulangan eksperimen yang sangat banyak dan menggunakan frekuensi relatif berbagai nilai X sebagai pendekatan nilai peluangnya.

Contoh 1.6

Misalkan X menunjukkan banyak jenis majalah yang biasa dibaca oleh seorang mahasiswa di kota A. Dari suatu survei dengan 400 orang mahasiswa di kota itu, diperoleh distribusi frekuensi sebagaimana tertuang dalam Tabel 1.4. Dengan memandang frekuensi relatif sebagai taksiran empiris nilai peluang, pada dasarnya kita telah memperoleh bentuk pendekatan distribusi peluang dari X. Distribusi peluang yang sebenarnya akan kita peroleh jika seluruh mahasiswa dalam populasi itu kita survei.

Tabel 1.4. Distribusi Frekuensi dari *X*

x	Frekuensi	Frekuensi Relatif
0	61	0,15
1	153	0,38
2	106	0,27
3	56	0,14
4	24	0,06
Jumlah	400	1,00

Harus kita tanamkan dalam pikiran kita perbedaan yang penting antara distribusi frekuensi relatif dan distribusi peluang. Distribusi frekuensi relatif adalah sesuatu yang timbul berdasarkan sampel sehingga dengan demikian memuat variasi pada sampel-sampel yang berbeda. Sebaliknya, distribusi peluang adalah sesuatu yang stabil yang menggambarkan seluruh populasinya. Ini adalah bentukan teoretis yang berguna sebagai model untuk menggambarkan variasi dalam populasinya.

Distribusi peluang dari X dapat digunakan untuk menghitung peluang peristiwa yang didefinisikan dalam bentuk X. Untuk menggambarkan hal ini, kita pandang distribusi peluang dalam Tabel 1.5 berikut.

Tabel 1.5.

x	$f(x)$
0	0,02
1	0,23
2	0,40
3	0,25
4	0,10

Peluang bahwa X lebih besar atau sama dengan 2 dapat dihitung sebagai $P(X \geq 2) \;= f(2) + f(3) + f(4)$

$$= 0,40 + 0,25 + 0,10 = 0,75$$

Dengan cara yang sama dapat juga kita hitung:

$$P(X \leq 2) \;= f(0) + f(1) + f(2)$$
$$= 0,02 + 0,23 + 0,40 = 0,65$$

Latihan 1

Untuk memperdalam pemahaman Anda mengenai materi di atas, silakan Anda mengerjakan latihan berikut ini!

1) Tabel di bawah ini menunjukkan hasil-hasil elementer suatu eksperimen dengan nilai peluang dan nilai variabel random X untuk tiap hasil. Hitunglah distribusi peluang dari X.

Hasil Elementer	**Peluang**	**Nilai *X***
e_1	0,12	2
e_2	0,29	0
e_3	0,05	0
e_4	0,08	3
e_5	0,16	1
e_6	0,11	1
e_7	0,09	2
e_8	0,10	3

2) Misalkan variabel random X menunjukkan banyaknya tampak belakang dalam tiga pelemparan sebuah mata uang, maka:
   a. hitunglah distribusi peluang dari X
   b. gambarkan histogram peluangnya.

3) Misalkan X menunjukkan banyak anak yang dipunyai satu keluarga di daerah KPR. Dari pemeriksaan 381 kartu keluarga yang dipilih secara random dari kantor pemerintah setempat, diperoleh distribusi frekuensi berikut. Hitunglah distribusi peluang pendekatan dari X.

Banyak Anak	**Banyak Keluarga**
0	2
1	82
2	161
3	89
4	47
Jumlah	381

Petunjuk Jawaban Latihan

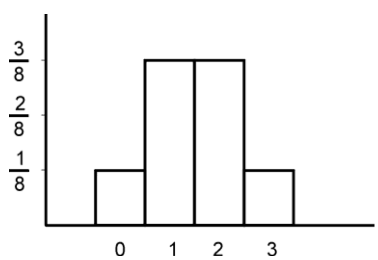
1) Distribusi peluang dari *X* adalah:

x	$f(x)$
0	0,34
1	0,27
2	0,21
3	0,18
Jumlah	1,00

2) *X* = banyak muka dalam tiga lemparan, maka:
   a. Distribusi peluang dari *X* adalah:

x	$f(x)$
0	1/8
1	3/8
2	3/8
3	3/8
Jumlah	1

   b. Histogram peluangnya adalah:



3) Distribusi peluang pendekatan dari *X* adalah sebagai berikut.

x	0	1	2	3	4	Jumlah
$f(x)$	$\frac{2}{381}$	$\frac{82}{381}$	$\frac{161}{381}$	$\frac{89}{381}$	$\frac{47}{381}$	1

Rangkuman

1. Hasil-hasil suatu eksperimen dikuantifikasi dengan menetapkan nilai angka yang berkaitan dengan sifat yang diinginkan. Aturan penetapan nilai angka itu dinamakan *variabel random.*
2. Distribusi peluang dari X menggambarkan bagaimana nilai peluang didistribusikan pada nilai-nilai X yang mungkin.
3. Distribusi peluang berguna sebagai model untuk menjelaskan variasi di dalam suatu populasi.

Tes Formatif 1

Pilih satu jawaban yang paling tepat dari beberapa alternatif jawaban yang disediakan.

1) Misalkan X menunjukkan banyak muka dalam empat pelemparan satu mata uang logam yang seimbang, maka banyaknya nilai X yang mungkin, adalah ….
   A. 2
   B. 3
   C. 4
   D. 5

2) **Lihat soal nomor 1**, nilai $P(X = 2)$ sama dengan ….
   A. 1/8
   B. 2/8
   C. 3/8
   D. 4/8

3) **Lihat soal nomor 1**, nilai $P(X \geq 2)$ sama dengan ….
   A. 10/16
   B. 11/16
   C. 12/16
   D. 13/16

4) Seorang pemain bola basket rata-rata dapat memasukkan bola sekali dalam tiga kali lemparan. Peluang bahwa dalam empat kali lemparan, dia tidak memasukkan bola sama sekali adalah ….
   A. 16/81
   B. 25/81
   C. 31/81
   D. 45/81

5) **Lihat soal nomor 4**, peluang bahwa dalam tiga kali lemparan, dia hanya memasukkan satu kali adalah ….
   A. 2/9
   B. 4/9
   C. 6/9
   D. 8/9

6) **Lihat soal nomor 4**, peluang bahwa dalam dua kali lemparan, dia memasukkan bola dua kali adalah ….
   A. 1/9
   B. 2/9
   C. 6/9
   D. 8/9

7) Dipunyai distribusi peluang sebagai berikut.

x	0	1	2	3	4	5
$f(x)$	0,05	0,20	k	0,25	0,15	0,05

Nilai k sama dengan ….
A. 0,15
B. 0,20
C. 0,25
D. 0,30

8) **Lihat soal nomor 7**, nilai $P(X \leq 2)$ sama dengan ….
A. 0,50
B. 0,55
C. 0,60
D. 0,65

9) Fungsi peluang suatu variabel random X diberikan dengan rumus:
$$f(x) = \frac{32}{31}\left(\frac{1}{2^x}\right);\ x = 1,\ 2,\ 3,\ 4,\ 5$$
Nilai $P(X = 3)$ sama dengan ….
A. 1/31
B. 2/31
C. 3/31
D. 4/31

10) **Lihat soal nomor 9**, nilai $P(X \geq 4)$ sama dengan ….
A. 1/31
B. 2/31
C. 3/31
D. 4/31

Setelah mengerjakan Tes Formatif 1 di atas, cocokkan jawaban Anda dengan Kunci Jawaban Tes Formatif 1 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar, kemudian gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 1.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{10} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:

90% - 100%	=	baik sekali
80% - 89%	=	baik
70% - 79%	=	cukup
< 70%	=	kurang

Apabila tingkat penguasaan Anda mencapai 80% atau lebih **Bagus!** Anda cukup memahami materi Kegiatan Belajar 1. Anda dapat meneruskan dengan Kegiatan Belajar 2. Tetapi apabila tingkat penguasaan masih di bawah 80%, Anda harus mengulangi Kegiatan Belajar 1, terutama bagian yang belum dikuasai.

Kegiatan Belajar

2

Beberapa Sifat Distribusi Peluang

Sekarang kita akan mengenalkan ukuran numerik untuk pusat suatu distribusi peluang dan juga untuk pencarannya. Mean merupakan ukuran pusat suatu himpunan data, dan deviasi standar sebagai ukuran pencarannya. Karena distribusi peluang adalah model teoretis dimana peluang dapat dipandang sebagai frekuensi relatif jangka panjang, maka ukuran pusat dan pencaran sampel mempunyai kawan imbangan untuk populasinya.

Untuk memotivasi definisi ukuran pusat dan pencaran populasi ini, marilah kita lihat bagaimana menghitung mean dan deviasi standar suatu himpunan data. Misalkan satu dadu dilemparkan 20 kali dan diperoleh hasil (titik-titik) sebagai berikut.

4	3	4	2	5	1	6	6	5	2
2	6	5	4	6	2	1	6	2	4

Mean observasi-observasi ini dinamakan *mean sampel* dan dihitung sebagai berikut.

$$\bar{x} = \frac{\text{jumlah semua observasi}}{\text{ukuran sampel}} = \frac{76}{20} = 3{,}8$$

Dengan cara lain, pertama-tama kita cacah frekuensi untuk tiap titik yang diperoleh dan menggunakan frekuensi relatifnya untuk menghitung mean itu, kita peroleh:

$$\bar{x} = 1\left(\frac{2}{20}\right) + 2\left(\frac{5}{20}\right) + 3\left(\frac{1}{20}\right) + 4\left(\frac{4}{20}\right) + 5\left(\frac{3}{20}\right) + 6\left(\frac{5}{30}\right) = 3{,}8$$

Cara hitungan yang kedua ini melukiskan rumus:

$$\bar{x} = \sum (\text{nilai} \times \text{frekuensi relatif})$$

Jika kita tidak hanya berhenti sampai 20 lemparan, melainkan kita lemparkan dadu itu sangat banyak kali, maka frekuensi relatifnya akan mendekati nilai peluangnya, yakni tiap titik berpeluang $\frac{1}{6}$ untuk sebuah dadu yang seimbang. Maka mean populasi lemparan (tak berhingga) sebuah dadu yang seimbang haruslah dihitung sebagai:

$$1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \ldots + 6\left(\frac{1}{6}\right) = 3{,}5$$

Dimotivasi oleh contoh tersebut dan stabilitas frekuensi relatif jangka panjang, maka wajarlah jika kita mendefinisikan mean suatu variabel random *X* atau mean distribusi peluangnya sebagai:

$$\text{mean} = \sum (\text{nilai} \times \text{peluang})$$

atau

$$\text{mean} = \sum x_i \; f(x_i)$$

di mana x_i menunjukkan nilai-nilai X yang berbeda-beda. Mean suatu distribusi peluang juga dinamakan mean populasi untuk variabel *X* dan ditulis dengan μ. Mean suatu variabel random *X* juga dinamakan *nilai harapan* dan sering juga ditulis dengan $E(X)$. Jadi, mean μ dan nilai harapan $E(X)$ adalah kuantitas yang sama dan akan digunakan bersama. Sehingga *mean X* atau *mean populasi* adalah:

$$
\begin{aligned}
E(X) &= \mu \\
&= \sum (\text{nilai} \times \text{peluang}) \\
&= \sum x_i \ f(x_i)
\end{aligned}
$$

di sini jumlah itu meliputi seluruh nilai x_i yang berbeda.

Contoh 1.7

Jika *X* menunjukkan banyak muka dalam tiga pelemparan sebuah mata uang logam, maka mean *X* dapat dihitung sebagai berikut.

Distribusi peluang *X* telah diberikan pada Tabel 1.1. Dari hitungan yang ditunjukkan dalam Tabel 1.6. berikut, kita peroleh mean sama dengan 1,5.

Tabel 1.6.

x	***f(x)***	***x f(x)***
0	1/8	0
1	3/8	3/8
2	3/8	6/8
3	1/8	3/8
Jumlah	1	$\mu = 12/8 = 1,5$

Mean suatu distribusi peluang mempunyai interpretasi fisik. Jika lempeng baja dipotong dalam bentuk histogram peluang, maka mean menunjukkan titik pada dasar histogram di mana lempeng baja itu seimbang. Misalnya mean $\mu = 1,5$ yang dihitung dalam Contoh 1.7 tepat terletak pada pusat massa untuk distribusi yang ditunjukkan dalam Gambar 1.1. Karena besar peluang berkaitan dengan banyak massa dalam persegi panjang, maka titik keseimbangan (titik berat) μ dapat kita interpretasikan sebagai pusat distribusi peluang.

Seperti banyak konsep peluang yang lain, dasar pemikiran mean atau nilai harapan berasal dari studi tentang perjudian. Jika *X* menunjukkan perolehan uang dalam permainan peluang seperti bermain poker misalnya, istilah *perolehan yang diharapkan* lebih menarik daripada *mean perolehan*. Dalam statistika, nama *mean* dan *nilai harapan* sama-sama digunakan secara luas.

Contoh 1.8

Suatu polis asuransi perjalanan akan membayar \$1.000 kepada pemegangnya jika terjadi kehilangan karena pencurian atau kerusakan dalam perjalanan lima hari. Jika risiko kehilangan seperti itu diperkirakan terjadi satu dalam dua ratus, maka premi yang adil untuk polis ini dapat dihitung sebagai berikut.

Peluang bahwa perusahaan asuransi akan dibebani membayar \$1.000 kepada pemegang polis adalah $1/200 = 0,005$. Dengan demikian distribusi peluang pembayaran per pemegang polis adalah sebagai berikut.

Pembayaran (*x*)	**Peluang *f*(*x*)**
\$ 0	0,995
\$ 1.000	0,005

Selanjutnya, $E(X) = (\$\ 0)(0,995) + (\$\ 1.000)(0,005) = \$\ 5$ maka biaya yang diharapkan adalah \$ 5 per pemegang polis dan premi sebesar jumlah uang ini dipandang sebagai premi yang adil. Jika premi ini ditetapkan dan tidak ada biaya lain yang terlibat, maka dalam jangka panjang perusahaan itu tidak akan memperoleh keuntungan atau akan menderita kerugian uang. Dalam praktik, nilai premi ditetapkan lebih tinggi, karena meliputi biaya administrasi dan keuntungan yang diinginkan.

Konsep nilai harapan juga membawa kita ke ukuran numerik yang lain, yakni ukuran numerik untuk pencaran suatu distribusi peluang yang dinamakan *deviasi standar*. Karena mean μ adalah pusat distribusi *X*, maka variasi *X* kita nyatakan dalam bentuk deviasi $(x - \mu)$. Variansi *X* kita definisikan sebagai nilai harapan deviasi kuadrat $(x - \mu)^2$. Untuk menghitung nilai harapan ini kita catat bahwa:

Nilai-nilai $(x - \mu)^2$	**Peluang**
$(x_1 - \mu)^2$	$f(x_1)$
$(x_2 - \mu)^2$	$f(x_2)$
$(x_3 - \mu)^2$	$f(x_3)$
$\vdots$	$\vdots$
$(x_k - \mu)^2$	$f(x_k)$

Nilai harapan $(X - \mu)^2$ diperoleh dengan mengalikan nilai-nilai $(x_i - \mu)^2$ dengan peluangnya $f(x_i)$ dan menjumlahkan hasil kalinya, maka diperoleh:

$$\begin{aligned}\operatorname{var}(X) &= \sum (deviasi)^2 \times (peluang) \\ &= \sum (x_i - \mu)^2 \; f(x_i) \\ &= E(X - \mu)^2\end{aligned}$$

Variansi X disingkat dengan $\operatorname{var}(X)$ dan juga ditulis dengan lambang σ^2. Deviasi standar X adalah akar positif dari $\operatorname{var}(X)$, ditulis dengan $sd(X)$ atau σ. Secara singkat dirumuskan:

$$\sigma^2 = \operatorname{var}(X) = E(X - \mu)^2$$

$$\sigma = sd\,(X) = \sqrt{\operatorname{var}(X)}$$

Contoh 1.9

Hitung variansi dan deviasi standar distribusi X yang ada dalam dua kolom paling kiri Tabel 1.7 berikut ini.

Tabel 1.7. Menghitung Variansi dan Deviasi Standar

x	$f(x)$	$x\,f(x)$	$(x-\mu)^2$	$(x-\mu)^2\,f(x)$
0	0,1	0	4	0,4
1	0,2	0,2	1	0,2
2	0,4	0,8	0	0,0
3	0,2	0,6	1	0,2
4	0,1	0,4	4	0,4
Jumlah	1,0	2,0		1,2

$\text{var}(X) = \sigma^2 = 1{,}2$ dan $sd(X) = \sqrt{\sigma^2} = \sqrt{1{,}2} = 1{,}095$

Sifat-sifat Mean dan Deviasi Standar

Misalkan X adalah variabel random dengan distribusi peluang $f(x)$. Andaikan kita tertarik pada variabel random baru $g(X)$, suatu fungsi X. Maka mean atau nilai harapan variabel random baru ini dapat kita hitung dengan dua cara, yaitu:

1. Kita hitung distribusi peluang $g(X)$ dari distribusi peluang X, selanjutnya nilai harapannya dihitung dengan cara seperti yang telah kita bicarakan.
2. Dihitung langsung dari distribusi peluang X. Untuk setiap nilai X kita hitung nilai $g(X)$, yang selanjutnya kita kalikan dengan peluang X. Jumlah hasil kali untuk semua nilai X adalah nilai harapan $g(X)$ atau $E[g(X)]$, jadi:

$$E[g(X)] = \sum g(x)\, f(x)$$

Contoh 1.10

Dipunyai distribusi peluang dari X seperti tertuang dalam dua kolom pertama Tabel 1.8 berikut. Jika $g(X) = 3X - 1$, maka nilai harapan $g(X)$ dapat dihitung sebagai berikut.

Tabel 1.8. Nilai Harapan

x	$f(x)$	$(3x-1)$	$(3x-1)\,f(x)$
0	0,1	-1	- 0,1
1	0,2	2	0,4
2	0,4	5	1,5
3	0,2	8	1,6
4	0,1	11	2,2
Jumlah			5,6

Jadi $E\left[g(X)\right]=E(3X-1)=5,6$

Selanjutnya akan kita pelajari sifat-sifat nilai harapan yang penting yang dapat mempermudah hitungan-hitungan nilai harapan.

1. Jika a dan b konstanta, maka $E(aX+b)=\ aE(X)+b$

   Bukti:

   Kebenaran sifat ini dapat kita lihat dari definisi nilai harapan $g(X)$, dalam hal ini $g(X)=aX+b$, sehingga:

   $$\begin{aligned} E(aX+b) &= \sum(ax+b)\,f(x) \\ &= \sum(ax)\,f(x)+\sum b\,f(x) \\ &= a\sum x\,f(x)+b\sum f(x)=aE(X)+b \end{aligned}$$

2. Jika kita ambil $a=0$, maka $E(b)=b$

3. Jika kita ambil $b=0$, maka $E(aX)=aE(X)$

4. Jika kita Nilai harapan jumlah atau selisih dua fungsi (atau lebih) sama dengan jumlah atau selisih nilai harapan masing-masing fungsi itu, yaitu:

   $$E\left[g(X)\pm h(X)\right]=E\left[g(X)\right]\pm E\left[h(X)\right]$$

Contoh 1.11

Marilah kita pandang kembali distribusi peluang dari *X* dalam Contoh 1.10. Nilai harapan *X*, kita peroleh sebagai berikut.

$$
\begin{aligned}
E(X) &= \sum x\, f(x) \\
&= 0(0,1) + 1(0,2) + 2(0,3) + 3(0,2) + 4(0,2) = 2,2
\end{aligned}
$$

$$
\begin{aligned}
E\left[g(X)\right] &= E(3X - 1) \\
&= 3E(X) - 1 \\
&= 3(2,2) - 1 = 5,6
\end{aligned}
$$

Selanjutnya, marilah kita pelajari sifat-sifat variansi dan deviasi standar *X*. Kita mulai dari definisi variansi *X*, yaitu:

$$
\begin{aligned}
\operatorname{var}(X) &= E(X - \mu)^2 \\
&= E\left(X^2 - 2\mu X + \mu^2\right) \\
&= E\left(X^2\right) - 2\mu E(X) + \mu^2 \\
&= E\left(X^2\right) - \mu^2
\end{aligned}
$$

Contoh 1.12

Dipunyai distribusi peluang *X* seperti tertuang dalam dua kolom pertama Tabel 1.9 berikut ini. Akan dihitung variansi *X* dan deviasi standar dari *X*.

Tabel 1.9. Menghitung Variansi *X*

x	$f(x)$	$x\,f(x)$	$x^2 f(x)$
0	0,1	0	0
1	0,1	0,1	0,1
2	0,2	0,4	0,8
3	0,3	0,9	2,7
4	0,2	0,8	3,2
5	0,1	0,5	2,5
Jumlah		2,7	9,3

$\mu = E(X) = 2,7$

$\text{var}(X) = E(X^2) - \mu^2 = 9,3 - (2,7)^2 = 2,01$

$sd(X) = \sqrt{\text{var}(X)} = \sqrt{2,01} = 1,42$

Karena variansi didefinisikan dalam bentuk nilai harapan, maka variansi (juga deviasi standar) juga mempunyai sifat-sifat yang dapat diturunkan dari sifat-sifat nilai harapan, yaitu:

1. $\text{var}(a) = E(a^2) - [E(a)]^2 = a^2 - a^2 = 0$
2. $\text{var}(X + a) = \text{var}(X)$

   Bukti:

   $$\begin{aligned}\text{var}(X + a) &= E[(X + a) - E(X + a)]^2 \\ &= E[(X) + a - E(X) - a]^2 \\ &= E[(X) - E(X)]^2 = \text{var}(X)\end{aligned}$$

3. $\text{var}(bX) = b^2 \text{var}(X)$

   Bukti:

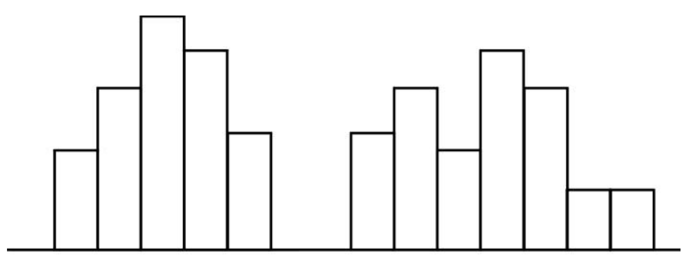
   $$\begin{aligned}\text{var}(bX) &= E(bX)^2 - [E(bX)]^2 \\ &= E(b^2 X^2) - [bE(X)]^2 \\ &= b^2 E(X^2) - b^2 [E(X)]^2 \\ &= b^2 \left\{E(X^2) - [E(X)]^2\right\} = b^2 \text{var}(X)\end{aligned}$$

4. $\text{var}(X + a) = \text{var}(X)$
5. $sd(X + a) = sd(X)$
6. $sd(bX) = b\,sd(X)$
7. $sd(bX + a) = b\,sd(X)$

Satu aplikasi penting sifat mean dan deviasi standar kita jumpai dalam transformasi variabel random X yang mempunyai mean μ dan deviasi standar σ menjadi variabel random Z yang mempunyai mean 0 dan deviasi standar 1. Variabel random Z dinamakan *variabel random unit standar*, transformasi ini adalah:

$$Z = \frac{X - \mu}{\sigma}$$

Arti mean dan deviasi standar suatu variabel random akan lebih jelas jika dikaitkan dengan satu teorema yang dikenal sebagai *pertidaksamaan Chebyshev*. Deviasi standar suatu variabel random mengukur pencaran atau deviasi setiap nilai X terhadap meannya. Karena itu jika suatu variabel random mempunyai deviasi standar kecil, dapat kita harapkan bahwa nilai-nilai variabel itu akan menggerombol dekat dengan meannya (μ). Sehingga peluang bahwa variabel random itu akan mempunyai nilai di dalam interval tertentu di sekitar mean akan lebih besar dari suatu variabel random lain yang mempunyai deviasi standar yang lebih besar. Jika kita pikirkan peluang dalam bentuk luasan histogram peluang, maka kita akan mengharapkan bahwa suatu distribusi peluang dengan deviasi standar kecil sebagian besar luasannya akan dekat dengan μ seperti terlihat dalam Gambar 1.3a . Sedangkan untuk distribusi yang deviasi standarnya lebih besar, histogram peluangnya akan lebih memencar seperti pada Gambar 1.3b.



(a) (b)

Gambar 1.3. Deviasi terhadap Mean

Seorang matematikawan bangsa Rusia, *Chebyshev*, telah menemukan bahwa bagian dari luasan di antara dua nilai simetrik terhadap mean berhubungan dengan nilai deviasi standarnya. Karena luasan dalam histogram peluang merupakan nilai peluang, maka luasan itu sama dengan peluang bahwa suatu variabel random akan bernilai antara dua nilai tersebut. Pertidaksamaan *Chebyshev* memberikan taksiran peluang bahwa suatu variabel random akan jatuh dalam interval k deviasi standar dari mean. Untuk sembarang konstan $k > 0$ berlaku:

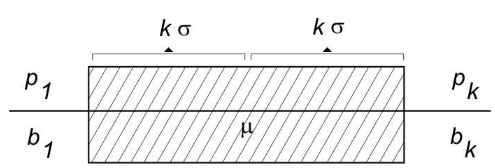
$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

atau

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

atau

$$P(|X - \mu| \geq k\sigma) < \frac{1}{k^2}$$



Gambar 1.4. Interpretasi Geometrik Pertidaksamaan *Chebyshev*

Pertidaksamaan *Chebyshev* di atas mempunyai interpretasi geometrik sebagai berikut. Gambarkan semua nilai b_i yang dapat dijalani oleh X pada suatu sumbu, masing-masing disertai oleh nilai peluang yang bersangkutan p_i (lihat Gambar 1.4). Dengan nilai μ sebagai pusat, kita gambarkan persegi panjang yang berjarak $k\sigma$ ke kiri dan ke kanan dari μ di mana k sebarang konstan positif. Jika eksperimen dilakukan, maka peluang bahwa nilai observasi X akan terletak di luar persegi panjang sama dengan jumlah p_i yang bersesuaian dengan nilai--nilai b_i yang terletak di luar empat persegi panjang itu. Peluang ini paling besar sama dengan $\frac{1}{k^2}$. Jika k makin besar, maka $\frac{1}{k^2}$ makin dekat dengan nol. Sebagai contoh, misalkan $\mu = 12$ dan $\sigma = 2$, maka dengan memilih $k = 3$, diperoleh pertidaksamaan *Chebyshev*:

$$P\left(\left|X - 12\right| \geq (3)(2)\right) = P\left(\{X \geq 18\} \text{ atau } \{X \leq 6\}\right) \leq \frac{1}{9} = 0{,}111$$

Sedangkan jika kita mengambil $k = 5$, kita peroleh:

$$P\left(\left|X - 12\right| \geq (5)(2)\right) = P\left(\{X \geq 22\} \text{ atau } \{X \leq 2\}\right) \leq \frac{1}{25} = 0{,}04$$

Dalam banyak hal pertidaksamaan *Chebyshev* agak konservatif, dalam arti bahwa ruas kiri akan jauh lebih kecil dari $\frac{1}{k^2}$. Pentingnya pertidaksamaan *Chebyshev* adalah akan memberi batas atas pada peluang yang sebenarnya peristiwa $\left\{\left|X - \mu\right| \geq k\sigma\right\}$.

✍ Latihan 2

Untuk memperdalam pemahaman Anda mengenai materi di atas, silakan Anda mengerjakan latihan berikut ini!

1) Dipunyai distribusi peluang sebagai berikut.

x	0	1	2
***f*(*x*)**	0,3	0,4	0,3

a. hitunglah mean μ
b. hitunglah variansi σ^2 dan deviasi standar σ
c. hitunglah mean *Y* dan deviasi standar *Y*, jika $Y = 4X - 2$
d. hitunglah mean *Z*, variansi *Z*, dan deviasi standar *Z*, jika $Z = 2X + 1$

2) Suatu perusahaan asuransi menjual polis yang melindungi rumahnya terhadap kebakaran untuk periode waktu 2 tahun. Berdasarkan biaya penggantian rumah yang terbakar sebesar 300 juta rupiah, perusahaan memperhitungkan akan tidak untung dan tidak rugi apabila peluang satu rumah milik pemegang polis akan terbakar selama periode polis, adalah sebesar 0,15. Dianggap bahwa peluang akan terjadi lebih dari satu kebakaran adalah nol.
   a. tentukan harga jual polis asuransi itu,
   b. jika peluang satu rumah akan terbakar hanya 0,10, hitunglah keuntungan tiap polis yang diharapkan perusahaan itu jika harga jual seperti dalam a.

Petunjuk Jawaban Latihan

1) Diketahui distribusi peluang sebagai berikut.

x	0	1	2
***f*(*x*)**	0,3	0,4	0,3

   a. mean $\mu = (0)(0,3) + (1)(0,4) + (2)(0,3) = 1$
   b. $E\left(X^2\right) = \left(0^2\right)(0,3) + \left(1^2\right)(0,4) + \left(2^2\right)(0,3) = 1,6$
      $\sigma^2 = E\left(X^2\right) - \left[E(X)\right]^2 = 0,6$ dan $\sigma = \sqrt{0,6} = 0,7746$
   c. $E(Y) = 4E(X) - 2 = 4(1) - 2 = 2$
      $sd(Y) = 4\,sd(X) = 4\sqrt{0,6} = 3,0984$
   d. $E(Z) = 2E(X) + 1 = 2(1) + 1 = 3$
      $sd(Z) = 2\,sd(X) = 2\sqrt{0,6} = 1,5492$

2) a. $E(X) = (-300)(0,15) + (x)(0,85) = 0$
      Jadi $x =$ Rp52,9412 juta
   b. $E(X) = (-300)(0,10) + (52,9412)(0,90) = 17,6471$

Rangkuman

1. Distribusi peluang mempunyai mean μ yang di interpretasikan sebagai *mean populasi*, kuantitas ini juga dinamakan nilai harapan $E(X)$.

$$\mu = E(X) = \sum (\text{nilai} \times \text{peluang}) = \sum x_i \; f(x_i)$$

2. Variansi populasi adalah:

$$\begin{aligned} \sigma^2 &= E(X-\mu)^2 \\ &= \sum (x-\mu)^2 \; f(x) \\ &= E(X)^2 - \mu^2 \end{aligned}$$

3. Deviasi standar populasi adalah akar positif variansi. Deviasi standar adalah ukuran pencaran atau variasi populasi.
4. Jika *X* variabel random dengan mean μ dan deviasi standar σ, maka *variabel random standar* adalah:

$$Z = \frac{X-\mu}{\sigma}$$

   Variabel random *Z* mempunyai nilai mean 0 dan deviasi standar 1.
5. Pertidaksamaan *Chebyshev* untuk sembarang konstan $k > 0$ berlaku:

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

   atau

$$P\left(|X-\mu| < k\sigma\right) > 1 - \frac{1}{k^2}.$$

☑ Tes Formatif 2

Pilih satu jawaban yang paling tepat dari beberapa alternatif jawaban yang disediakan.

1) Dipunyai distribusi peluang dari *X* sebagai berikut:

x	-2	-1	0	1	2
***f*(*x*)**	0,1	0,2	0,3	0,3	0,1

Maka mean *X* sama dengan ….
A. 0,0
B. 0,1
C. 0,2
D. 0,3

2) **Lihat soal nomor 1**, deviasi standar *X* sama dengan ….
A. 1,1
B. 1,2
C. 1,3
D. 1,4

3) **Lihat soal nomor 1**, jika $Y = 2X + 3$, maka mean *Y* sama dengan ….
A. 2,3
B. 3,2
C. 4,1
D. 5,4

4) **Lihat soal nomor 3**, maka variansi *Y* sama dengan ….
A. 3,16
B. 4,16
C. 5,16
D. 6,16

5) **Lihat soal nomor 3**, maka deviasi standar Y sama dengan ….
   A. 2,27
   B. 3,27
   C. 4,27
   D. 5,27

6) Seorang pengacara merasa bahwa peluang dia akan memenangkan suatu perkara adalah 0,3. Jika dia menang, maka dia akan memperoleh Rp30.000, tetapi jika dia kalah, dia tidak akan mendapatkan apa pun. Nilai harapan pengacara akan memperoleh kemenangan adalah ….
   A. Rp900
   B. Rp9.000
   C. Rp90.000
   D. Rp900.000

7) **Lihat soal nomor 6**, dalam mempersiapkan perkara itu, pengacara menghabiskan biaya Rp5.000, harapan perolehan bersih pengacara itu adalah ….
   A. Rp1.000
   B. Rp2.000
   C. Rp3.000
   D. Rp4.000

8) Misalkan X variabel random diskret dengan mean 10 dan variansi 9, maka batas atas $P\left(\left|X-10\right|\geq 6\right)$ adalah ….
   A. 1/2
   B. 1/3
   C. 1/4
   D. 1/5

9) **Lihat soal nomor 8**, batas atas $P\left(\left|X-10\right|\geq 9\right)$ adalah ….
   A. 1/4
   B. 1/9
   C. 1/16
   D. 1/25

10) **Lihat soal nomor 8**, batas bawah $P\left(\left|X-10\right|\leq 12\right)$ adalah ….

A. 1/16
B. 9/16
C. 12/16
D. 15/16

Setelah mengerjakan Tes Formatif 2 di atas, cocokkan jawaban Anda dengan Kunci Jawaban Tes Formatif 2 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar, kemudian gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 2.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{10} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:
90% - 100% = baik sekali
80% - 89% = baik
70% - 79% = cukup
< 70% = kurang

Apabila tingkat penguasaan Anda mencapai 80% ke atas **Bagus!** Anda cukup memahami materi Kegiatan Belajar 2. Anda dapat meneruskan dengan Kegiatan Belajar 3. Tetapi bila tingkat penguasaan masih di bawah 80%, Anda harus mengulangi Kegiatan Belajar 2, terutama bagian yang belum dikuasai.

Kegiatan Belajar

3

Distribusi Peluang Bersama

Sebagai hasil suatu eksperimen, sering kali kita mengamati lebih dari satu variabel random secara bersama-sama. Dalam hal ini kita tidak hanya ingin mempelajari distribusi peluang masing-masing variabel random, melainkan kita juga akan mempelajari distribusi peluang bersama beberapa variabel random yang kita amati. Dalam kesempatan ini kita hanya akan mempelajari dua variabel random.

Misalkan *X* dan *Y* dua variabel random, distribusi peluang akan terjadinya *X* dan *Y* secara bersama-sama dapat ditulis dengan notasi fungsional $f(x, y)$. Fungsi $f(x, y)$ ini biasanya dinamakan distribusi peluang bersama *X* dan *Y*, sehingga:

$$f(x, y) = P(X = x, Y = y)$$

Misalnya suatu pesawat TV harus diperbaiki, *X* menunjukkan umur pesawat TV dan *Y* menunjukkan banyak tabung yang rusak dalam pesawat itu, maka $f(5, 3)$ adalah peluang bahwa pesawat TV itu berumur 5 tahun dan memerlukan 3 tabung yang baru.

Distribusi peluang bersama variabel random diskret *X* dan *Y* mempunyai sifat:

$$f(x, y) \geq 0 \text{ untuk semua } (x, y)$$

dan

$$\sum_x \sum_y f(x,y) = 1$$

Contoh 4.13

Dua bola diambil secara random dari sebuah kotak yang berisi empat bola merah, tiga bola biru, dan dua bola hijau. Jika X menyatakan banyak bola merah yang terambil dan Y banyak bola biru terambil, maka distribusi peluang bersama X dan Y dapat diperoleh sebagai berikut.

Pasangan-pasangan nilai (x, y) yang mungkin terjadi adalah $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, $(0,2)$ dan $(2,0)$. Nilai $(0,0)$ menunjukkan yang terambil keduanya adalah bola hijau, nilai $(1,0)$ menunjukkan satu bola merah dan satu bola hijau terambil, dan seterusnya. Banyak cara yang mungkin untuk mengambil dua bola dari sembilan bola yang ada adalah $\binom{9}{2} = 36$. Banyak cara untuk mendapatkan satu dari empat bola merah dan satu dari dua bola hijau adalah $\binom{4}{1}\binom{2}{1} = 8$. Jadi $f(1,0) = \frac{8}{36}$.

Dengan jalan pikiran yang sama, kita peroleh:

$$f(0,0) = \frac{\binom{2}{2}}{36} = \frac{1}{36}; \quad f(0,1)\frac{\binom{3}{1}\binom{2}{1}}{36} = \frac{6}{36}$$

dan seterusnya. Nilai peluang itu dapat kita sajikan dalam tabel di bawah ini.

Tabel 1.10. Distribusi Peluang Bersama

Y	***x***			**Jumlah**
	0	**1**	**2**	
0	1/36	8/36	6/36	15/36
1	6/36	12/36	0	18/36
2	3/36	0	0	3/36
Jumlah	10/36	20/36	6/36	1

Distribusi peluang bersama (X,Y) dalam Tabel 1.10 dapat juga kita sajikan dalam bentuk rumus fungsi peluang bersama sebagai berikut.

$$f(x,y)=\frac{\binom{4}{x}\binom{3}{y}\binom{2}{2-x-y}}{\binom{9}{2}}\;; \qquad \begin{array}{l} x=0,1,2;\; y=0,1,2\;; \\ 0\le x+y\le 2 \end{array}$$

Selanjutnya, kita hitung berbagai peluang, misalnya:

$$\begin{aligned} P(X+Y\le 1) &= f(0,0)+f(0,1)+f(1,0) \\ &= \frac{1}{36}+\frac{6}{36}+\frac{8}{36}=\frac{15}{36} \end{aligned}$$

Sekarang kita pelajari beberapa sifat fungsi peluang bersama (X,Y) sebagai berikut. Dipunyai distribusi peluang bersama X dan Y sebagai $f(x,y)$, distribusi peluang variabel random X sendiri dan juga Y sendiri masing-masing adalah sebagai berikut.

$$g(x)=\sum_y f(x,y)$$

$$h(y)=\sum_x f(x,y)$$

$g(x)$ dinamakan *distribusi peluang marginal* X dan $h(y)$ dinamakan *distribusi peluang marginal* Y. Dari distribusi peluang bersama , diperoleh Tabel 1.11 berikut.

Tabel 1.11.

$\boldsymbol{x}$	0	1	2	Jumlah
$\boldsymbol{g(x)}$	10/36	20/36	6/36	1

$\boldsymbol{y}$	0	1	2	Jumlah
$\boldsymbol{h(y)}$	15/36	18/36	3/36	1

Menggunakan definisi peluang bersyarat, yakni:

$$P(A|B)=\frac{P(A\cap B)}{P(B)}\ ;\ P(B)\neq 0$$

Kita dapat mengganti peristiwa *A* dan *B* masing-masing dengan peristiwa $\{X=x\}$ dan $\{Y=y\}$, sehingga kita peroleh:

$$P(X=x|\ Y=y)=\frac{P(X=x,\ Y=y)}{P(Y=y)}$$

$$=\frac{f(x,y)}{h(y)}\quad ;\quad h(y)\neq 0$$

Jika distribusi peluang itu ditulis dengan $f(x|y)$, maka kita punyai:

$$f(x|y)=\frac{f(x,y)}{h(y)}\ ;\ \ h(y)\neq 0$$

yang dinamakan *distribusi peluang bersyarat X*, jika diketahui $Y=y$. Demikian juga kita mendefinisikan $f(y|x)$ sebagai *distribusi peluang bersyarat Y*, jika diketahui $X=x$, kita tulis sebagai berikut.

$$f(y|x)=\frac{f(x,y)}{g(x)}\ ;\ \ g(x)\neq 0$$

Menggunakan contoh $f(x,y)$ dalam Tabel 1.10, kita dapat menghitung $f(x|1)=P(X=x|Y=1)$ sebagai berikut.

$$h(1)=\sum_{x=0}^{2} f(x,1)=f(0,1)+f(1,1)+f(2,1)$$

$$=\frac{6}{36}+\frac{12}{36}+0=\frac{18}{36}$$

Selanjutnya,

$$f(x|1)=\frac{f(x,1)}{h(1)}=\frac{f(x,1)}{18/36}$$

maka:

$$f(0|1)=\frac{6/36}{18/36}=\frac{6}{16}=\frac{1}{3}$$

$$f(1|1)=\frac{12/36}{18/36}=\frac{12}{18}=\frac{2}{3}$$

$$f(2|1)=\frac{0}{18/36}=0$$

dan distribusi peluang bersyarat dari $f(x|1)$ dapat disajikan dalam Tabel 1.12 berikut.

Tabel 1.12.

X	0	1	2	Jumlah
$\boldsymbol{f(x|1)}$	1/3	2/3	0	1

Variabel random *X* dan *Y* dikatakan *independen* jika dan hanya jika:

$$f(x,y)=g(x)\ h(y)\ \text{ untuk semua } (x,y)$$

Dalam contoh tersebut, kita peroleh:

$$f(1,0)=\frac{8}{36}$$

$$g(1)=\sum_{y=0}^{3}f(1,y)=\frac{8}{36}+\frac{12}{36}+0=\frac{20}{36}=\frac{5}{9}$$

$$h(0)=\sum_{x=0}^{3}f(x,0)=\frac{1}{36}+\frac{8}{36}+\frac{6}{36}=\frac{15}{36}=\frac{5}{12}$$

Karena $f(1,0)\neq g(1)\ h(0)$, maka *X* dan *Y* tidak independen.

Nilai Harapan dan Sifat-Sifatnya

Misalkan X dan Y variabel random dengan distribusi peluang bersama $f(x,y)$, maka nilai harapan suatu fungsi $g(X,Y)$ didefinisikan sebagai berikut.

$$E\left[g(X,Y)\right]=\sum_{x}\sum_{y}g(x,y)f(x,y)$$

Contoh 1.14

Misalkan X dan Y variabel random dengan distribusi peluang bersama seperti diberikan dalam Tabel 1.10, maka nilai harapan $g(X,Y)=XY$ dapat dihitung sebagai berikut.

$$\begin{aligned}E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2}xy\ f(x,y)\\ &=(0)(0)\,f(0,0)+(0)(1)\,f(0,1)+\ldots+(2)(0)\,f(2,0)\\ &= f(1,1)=\frac{12}{36}=\frac{1}{3}\end{aligned}$$

Perhatikan bahwa jika $g(X,Y)=X$ dan $h(X,Y)=Y$, maka:

$$E(X)=\sum_{x}\sum_{y}x\ f(x,y)=\sum_{x}x\sum_{y}f(x,y)=\sum_{x}x\ g(x)$$

$$E(Y)=\sum_{x}\sum_{y}x\ f(x,y)=\sum_{y}y\sum_{x}f(x,y)=\sum_{y}y\ h(y)$$

di mana $g(x)$ adalah distribusi peluang marginal X dan $h(y)$ adalah distribusi peluang marginal Y. Nilai harapan jumlah atau selisih dua fungsi variabel random X dan Y sama dengan jumlah atau selisih nilai harapan masing-masing, yakni:

$$E\left[g(X,Y)\pm h(X,Y)\right]=E\left[g(X,Y)\right]\pm E\left[h(X,Y)\right]$$

Jika $g(X,Y)=X$ dan $h(X,Y)=Y$, maka berlaku:

$$E(X \pm Y)=E(X) \pm E(Y)$$

Jika X dan Y dua variabel *independen,* maka berlaku:

$$E(X\,Y)=E(X)\,E(Y)$$

Jika $\mu_X = E(X)$ dan $\mu_Y = E(Y)$, maka *kovariansi* dua variabel random X dan Y didefinisikan sebagai:

$$\begin{aligned} kov(X,Y) &= E(X-\mu_X)(Y-\mu_Y) \\ &= E(X)E(Y)-\mu_X\,\mu_Y \end{aligned}$$

Kovariansi X dan Y sering ditulis dengan lambang σ_{XY}. Kovariansi akan *positif* jika nilai-nilai X yang tinggi berkaitan dengan nilai-nilai Y yang tinggi, dan nilai-nilai X yang rendah berkaitan dengan nilai-nilai Y yang rendah pula. Jika nilai-nilai X yang rendah berkaitan dengan nilai-nilai Y yang tinggi dan sebaliknya, maka kovariansinya akan *negatif*. Jika X dan Y independen, maka dapat ditunjukkan bahwa kovariansinya nol. Tetapi kebalikannya tidak selalu benar, yakni dua variabel random yang mempunyai kovariansi nol, tetapi tidak independen. Selanjutnya, jika a dan b sembarang konstanta, dapat ditunjukkan bahwa untuk dua variabel random X dan Y berlaku:

$$\mathrm{var}(aX+bY)=a^2\,\mathrm{var}(X)+b^2\,\mathrm{var}(Y)+2ab\,kov(X,Y)$$

$$\mathrm{var}(aX-bY)=a^2\,\mathrm{var}(X)+b^2\,\mathrm{var}(Y)-2ab\,kov(X,Y)$$

Jika X dan Y independen, maka $kov(X,Y)=0$, sehingga untuk kasus ini berlaku:

$$\mathrm{var}(aX \pm bY)=a^2\,\mathrm{var}(X)+b^2\,\mathrm{var}(Y)$$

Koefisien korelasi antara X dan Y didefinisikan sebagai:

$$kor(X,Y)=\frac{kov(X,Y)}{\sqrt{\mathrm{var}(X)\;\mathrm{var}(Y)}}=\frac{\sigma_{XY}}{\sigma_X\;\sigma_Y}$$

Koefisien korelasi adalah selalu bilangan antara –1 dan 1. Nilai-nilai ekstrim 1 dan –1 dicapai apabila X dan Y dihubungkan dengan garis lurus, masing-masing dengan lerengan positif dan negatif. Koefisien korelasi tetap tidak berubah jika konstanta ditambahkan pada variabel-variabel itu, atau jika variabel-variabel itu dikalikan dengan konstanta yang mempunyai tanda aljabar yang sama. Misalnya, jika $U = 3X + 1$ dan $V = 5Y + 4$, maka $kor(U,V) = kor(X,Y)$

✍ Latihan 3

Untuk memperdalam pemahaman Anda mengenai materi di atas, silakan Anda mengerjakan latihan berikut ini!

1) Diketahui distribusi peluang bersama X dan Y sebagai berikut.

y	***x***		
	0	**1**	**2**
0	0,1	0,3	0,1
1	0,2	0,2	0,1

Hitunglah:

a. $P(X = Y)$

b. $P(X > Y)$

c. Distribusi $(X + Y)$

2) Dipunyai distribusi peluang bersama (X,Y) sebagai berikut.

y	***x***	
	2	**4**
1	0,10	0,15
2	0,20	0,30
3	0,10	0,15

a. tentukan distribusi peluang marginal X,
b. tentukan distribusi peluang marginal Y,
c. tunjukkan bahwa X dan Y independen.

3) Jika X dan Y variabel random independen dengan $\text{var}(X) = 5$, $\text{var}(Y) = 3$, dan $Z = -2X + 4Y - 3$, maka hitunglah $\text{var}(Z)$.

Petunjuk Jawaban Latihan

1) a. $P(X = Y) = p(0,0) + p(1,1) = 0,1 + 0,2 = 0,3$
   b. $P(X > Y) = p(1,0) + p(2,0) + (2,1) = 0,3 + 0,1 + 0,1 = 0,5$
   c. Misalkan $Z = X + Y$, maka distribusi Z adalah:

z	0	1	2	3
f(z)	0,1	0,5	0,3	0,1

2) a. Distribusi peluang marginal *X* adalah:

x	2	4
f(x)	0,4	0,6

   b. Distribusi peluang marginal *Y* adalah:

y	1	2	3
f(y)	0,25	0,5	0,25

   c. Karena $f(x, y) = f(x)f(y)$ untuk setiap (x, y) maka X dan Y independen.

3) $\text{var}(Z) = 4\ \text{var}(X) + 16\ \text{var}(Y) = 4(5) + 16(3) = 68$

Rangkuman

1. Distribusi peluang bersama *X* dan *Y* ditulis sebagai:

$$f(x, y) = P(X = x, Y = y)$$

2. Distribusi peluang marginal dari *X* dan *Y* masing-masing adalah:

$$g(x) = \sum_y f(x, y) \text{ dan } h(y) = \sum_x f(x, y)$$

3. Distribusi peluang bersyarat dari *X*, jika $Y = y$ diketahui adalah:

$$f(x \mid y) = \frac{f(x, y)}{h(y)}$$

4. Distribusi peluang bersyarat dari *Y*, jika $X = x$ adalah:

$$f(y \mid x) = \frac{f(x, y)}{g(x)}$$

5. *X* dan *Y* independen, jika dan hanya jika:

$$f(x, y) = g(x)\, h(y) \text{ untuk semua } (x, y)$$

6. *X* dan *Y* independen, jika dan hanya jika:

$$E(X, Y) = E(X)E(Y)$$

7. Kovariansi *X* dan *Y* didefinisikan sebagai:

$$kov(X, Y) = E(X - \mu_X)(Y - \mu_Y) = E(XY) - \mu_X\, \mu_Y$$

8. Koefisien korelasi *X* dan *Y* adalah:

$$kor(X, Y) = \frac{kov(X, Y)}{\sqrt{\text{var}(X)\ \text{var}(Y)}}$$

☑ Tes Formatif 3

Pilih satu jawaban yang paling tepat dari beberapa alternatif jawaban yang disediakan.

1) Variabel random X dan Y mempunyai distribusi peluang bersama:

y	***x***		
	1	**2**	**3**
1	0	1/6	1/12
2	1/5	1/9	0
3	2/15	1/4	1/18

$E(X)$ sama dengan ….

A. 1,0
B. 1,4
C. 1,8
D. 2,2

2) **Lihat soal nomor 1**, nilai $\text{var}(X)$ sama dengan ….

A. 0,22
B. 0,33
C. 0,44
D. 0,5

3) **Lihat soal nomor 1**, nilai $E(Y)$ sama dengan ….

A. 1,8
B. 2,2
C. 2,5
D. 2,8

4) **Lihat soal nomor 1**, nilai $\text{var}(Y)$ sama dengan ….
   A. 0,61
   B. 0,72
   C. 0,79
   D. 0,83

5) **Lihat soal nomor 1**, nilai $kov(X,Y)$ sama dengan ….
   A. –0,47
   B. –0,13
   C. 0,56
   D. 1,32

6) **Lihat soal nomor 1**, nilai $kor(X,Y)$ sama dengan ….
   A. –0,11
   B. 0,18
   C. 0,36
   D. –0,25

7) **Lihat soal nomor 1**, nilai $P(X \geq Y)$ sama dengan ….
   A. 2/12
   B. 3/12
   C. 4/12
   D. 5/12

8) Jika diketahui $\text{var}(X)=8$, $\text{var}(Y)=18$ dan $kov(X,Y)=-10$ maka $kor(X,Y)$ sama dengan ….
   A. –0,83
   B. –0,73
   C. 0,53
   D. 0,63

9) **Lihat soal nomor 8**, nilai $\text{var}(5X-2Y)$ sama dengan ….
   A. 2272
   B. 2171

C. 2070
D. 1969

10) **Lihat soal nomor 8**, nilai $kor(3X+2;2Y-1)$ sama dengan ….
A. 2272
B. 2171
C. 2070
D. 1969

Setelah mengerjakan Tes Formatif 3 di atas, cocokkan jawaban Anda dengan Kunci Jawaban Tes Formatif 3 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar, kemudian gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 3.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{10} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:
90% - 100% = baik sekali
80% - 89% = baik
70% - 79% = cukup
< 70% = kurang

Apabila tingkat penguasaan Anda mencapai 80% ke atas **Bagus!** Anda cukup memahami materi Kegiatan Belajar 3. Anda dapat meneruskan dengan modul selanjutnya. Tetapi bila tingkat penguasaan masih di bawah 80%, Anda harus mengulangi Kegiatan Belajar 3, terutama bagian yang belum dikuasai.

Kunci Jawaban Tes Formatif

Tes Formatif 1

1)	D	6)	A
2)	C	7)	D
3)	B	8)	B
4)	A	9)	D
5)	B	10)	C

Tes Formatif 2

1)	B	6)	B
2)	A	7)	D
3)	C	8)	C
4)	C	9)	B
5)	A	10)	D

Tes Formatif 3

1)	C	6)	D
2)	C	7)	D
3)	B	8)	A
4)	A	9)	A
5)	B	10)	A

Daftar Pustaka

Bhattacharyya, G.K. and Johnson, R.A. (1977). *Statistical Concepts and Methods*. New York: Willey.

Freund, J. (1979). *Modern Elementary Statistics*. Prentice Hall.

Pfeffenberger, R.C. and Petterson, J.H. (1975). *Statistical Methods for Business and Economics*. Illions: Irwin.

Robbins, H. and Van Ryzin, J. (1975). *Introduction to Statistics*. Science Research Associates, Inc.